Review TRENDS in Ecology and Evolution Vol.22 No.5

Microarray challenges in ecology

Jan E. Kammenga[1], Michael A. Herman[2], N. Joop Ouborg[3], Loretta Johnson[2] and Rainer Breitling[4]

[1] Laboratory of Nematology, Wageningen University, Binnenhaven 5, 6709 PD, Wageningen, the Netherlands
[2] Division of Biology, Kansas State University, Manhattan, KS 66506-4901, USA
[3] Department of Ecology, Radboud University Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, the Netherlands
[4] Groningen Bioinformatics Centre, University of Groningen, Kerklaan 30, 9751 NN Haren, the Netherlands

Microarrays are used to measure simultaneously the amount of mRNAs transcribed from many genes. They were originally designed for gene expression profiling in relatively simple biological systems, such as cell lines and model systems under constant laboratory conditions. This poses a challenge to ecologists who increasingly want to use microarrays to unravel the genetic mechanisms underlying complex interactions among organisms and between organisms and their environment. Here, we discuss typical experimental and statistical problems that arise when analyzing genome-wide expression profiles in an ecological context. We show that experimental design and environmental confounders greatly influence the identification of candidate genes in ecological microarray studies, and that following several simple recommendations could facilitate the analysis of microarray data in ecological settings.

Microarrays in a variable environment

The use of gene expression microarrays (see Glossary) in ecology has increased rapidly over the past few years. They have been applied with the aim of understanding the genetic mechanisms that underlie species interactions, adaptation and the outcomes of evolutionary processes. Most of these studies were performed under laboratory or carefully controlled conditions (Table 1) using model species such as the fruit fly *Drosophila melanogaster,* thale cress *Arabidopsis thaliana* and baker's yeast *Saccharomyces cerevisiae*, which are well suited to laboratory experimentation and have fully sequenced genomes [1–3]. However, because model species have relatively simple life cycles and are opportunistic generalists, which limits their potential for ecological research, the number of field studies with non-model species is increasing.

However, detection of subtle gene effects in field studies might be hampered owing to large environmental variation. Therefore, microarray field experiments have focused mainly on differential gene expression associated with relatively large and discrete effects, such as dwarfism in fish [4], parasitism in birds [5,6] rearing conditions in salmon [7] and behavioral transitions in bees [8]. Yet, minimizing environmental variation has its limits because ecologists are interested in the interactions among organisms and between organisms and environmental heterogeneity. Here, we address the experimental and statistical caveats involved in linking high-throughput gene expression profiling using microarrays to ecological questions. Specific attention is paid to interpreting the results from ecological microarray studies.

Sources of variation

The analysis of microarray data is faced with many confounding factors. Whitehead and Crawford [9] mention three levels of variation: technical variation, variation among individuals, and variation between taxa. In addition, microarray analyses suffer from technical variation among platforms and laboratories [10]. Gene transcript abundance is also sensitive to a range of internal and external environmental variables, as illustrated for *D. melanogaster* by Carsten *et al.* [11], who showed that a simple dietary shift

Glossary

Anonymous array: a microarray in which the sequence of each probe is not known in advance. Only probes that show interesting expression changes in a particular experiment are sequenced. This strategy is particularly useful for species for which no genome data are available. Depending on the organism, anonymous probes can be derived from genomic DNA fragments or from libraries of mRNA (cDNA microarray).
cDNA microarray: a microarray in which probes are derived from expressed sequences by reverse transcribing mRNA. Probes are long (up to several hundred base pairs) and non-uniform, which can reduce the quality of signals compared with oligo microarrays.
Dedicated array: a microarray spotted with DNA probes of genes that are known to be involved in a particular pathway or underlie a specific phenotype (e.g. plant–herbivore interactions).
Effective population size: the number of breeding individuals in an idealized population that would show the same amount of random genetic drift and inbreeding as the population under study. This is usually smaller than the absolute population size.
Gene regulatory sequence: a DNA sequence that is responsible for regulating gene expression.
Microarray: a thumbnail-size sheet of glass or silicon on which thousands of single-stranded DNA probes are spotted. These sequences are complementary to pieces of genes of a particular biological species. When an mRNA sample from the same species is labeled with a fluorescent dye and applied to the array, it binds (i.e. hybridizes) to those probes that contain matching DNA. The arrays are then scanned and the amount of mRNA quantified. The fluorescent signal corresponds to the gene expression levels in the original sample and shows which genes are 'turned on'.
Northern blotting: a method of RNA detection and identification in which the intact RNA is separated by size, transferred (blotted) onto nitrocellulose or nylon paper, and then hybridized with labeled DNA probes.
Oligo microarray: a microarray that uses probes comprising synthesized pieces of DNA of uniform length (~ 40–80 base pairs). This can lead to more comparable signals than acquired using cDNA microarrays, but the technique is usually only used for organisms with completely sequenced genomes.
qRT-PCR: a technique that is used in combination with reverse-transcription PCR to quantify small amounts of mRNA in a sample. It is a popular method for validating microarray results for single genes.
Transcriptomics: the comprehensive measurement of mRNA levels (gene expression) in a particular biological sample, usually using microarrays.

Corresponding author: Kammenga, J.E. (Jan.Kammenga@wur.nl).
Available online 12 February 2007.

 0169-5347/$ – see front matter doi:10.1016/j.tree.2007.01.013

Table 1. Examples of studies using gene expression microarrays in an ecological context

Species	Array type	Subject	Cross-species array	Technical confirmation	Field/lab	Refs
Mycorrhizal fungus *Paxillus involutus*	cDNA microarray	Evolution of gene expression	No	No	Lab	[52]
		Host specificity	No	No	Lab	[53]
Baker's yeast *Saccharomyces cerevisiae*	cDNA microarray	Genetic variation in gene expression	No	No	Lab	[54]
		Adaptive evolution	No	No	Lab	[55]
Black cottonwood *Populus trichocarpa*	cDNA microarray	Plant defense	No	RT-PCR	Semi-field	[6]
Thale cress *Arabidopsis* spp.	cDNA microarray	Plant defense responses	No	No	Lab	[56]
	Oligo microarray	Adaptation, zinc accumulation	Yes	Yes	Lab	[57]
		Abiotic stress response	No	No	Lab	[58]
		Genetic variation of gene expression	No	RT-PCR	Lab	[48]
Sunflower *Helianthus* spp.	EST-anonymous	Habitat divergence	Yes	RT-PCR	Lab	[36]
Ragwort *Senecio* spp.	cDNA microarray	Hybrid speciation	Yes	RT-PCR	Lab	[59]
Tobacco *Nicotiana attenuata*	Dedicated cDNA microarray	Species interaction	No	No	Lab	[60]
	Dedicated cDNA array	Species interactions	No	No	Lab	[61]
		Plant defense responses	No	No	Lab	[31]
Black nightshade *Solanum nigrum*	Oligo microarray	Species interaction	No	RT-PCR, RNA blot analysis	Lab	[62]
Fruit fly *Drosophila melanogaster*	cDNA microarray, EST	Diet effects	No	No	Lab	[11]
	Oligo microarray	Species interactions	No	No	Lab	[63]
	Genome arrays	Life-history tradeoffs	No	No	Lab	[29]
		Body size	No	No	Lab	[64]
Honey bee *Apis mellifera*	cDNA microarray	Behavior	No	RT-PCR	Field	[8]
		Pheromone responses	No	RT-PCR	Lab	[65]
Annual killifish *Austrofundulus limnaeus*	Whole-genome cDNA array	Temperature acclimation	No	No	Lab	[66]
Lake whitefish *Coregonus clupeaformis*	cDNA microarray	Parallel transcription among ecotypes	Yes	No	Field	[4]
Mummychog *Fundulus heteroclitus*	cDNA microarray	Variation in cardiac gene expression	No	No	Lab	[67]
		Variation in tissue-specific gene expression	No	No	Lab	[68]
Salmon *Salmo salar*	cDNA microarray	Gene x environment interactions	Yes	No	Field	[7]
House finch *Carpodacus mexicanus*	cDNA macroarray	Parasite-induced gene expression	No	Northern blot hybridization	Semi-field	[5]

from cornmeal to banana for 24 hours was sufficient to trigger differential expression for 90 genes, which is nearly 2% of all genes on the microarray. Even in inbred mice, reared under highly controlled, pathogen-free laboratory conditions and matched for age and sex, significant inter-individual variation in gene expression was observed [12]. Only after controlling for additional variables such as social status, stress and food intake was the variation reduced. Comparable results were reported for the yeast transcriptome [13], where 4% of genes were differentially expressed between cultures of isogenic lines grown under identical laboratory conditions.

These variations showed consistent and biologically meaningful correlation patterns between groups of genes, indicating that they are not the result of technical noise but are instead caused by uncontrolled environmental factors. Even expression patterns that are highly constrained by evolution can be sensitive to environmental and physiological conditions. Ribosomal genes provide a striking example of precisely controlled protein synthesis, where unfavorable, or even slightly stressful conditions, lead to a rapid shutdown. Examples of this were reported for yeast following a metabolic shift from fermentation to respiration [14], nocturnal leaf growth in poplar [15], cold adaptation in catfish [16] and ultraviolet radiation acclimation in maize [17]. Thus, when microarrays are used to unravel the ecological complexity in field conditions, environmental variation can make the results difficult to interpret.

Stabilizing selection and neutral genetic drift

One of the implicit assumptions of ecological microarray studies is that expression levels are subject to evolutionary pressure and that intertaxa differences in expression are due to adaptation to different environments. Whitehead and Crawford [18] analyzed the expression of metabolic genes in populations of the fish *Fundulus heteroclitus* and found that many genes have expression patterns that cannot be explained by drift and show evidence for being under natural selection. This agrees with earlier conclusions that gene expression variation is largely determined by natural selection within and among species [19]. Changes in gene regulatory sequences can have large effects on gene expression. Wray *et al.* [20] reviewed the evolution of gene expression regulation in eukaryotes and reported an extensive genetic variation in regulatory sequences, some of which could be attributed to selection. Based on a meta-analysis of gene expression studies, Gilad *et al.* [21] concluded that stabilizing selection is the dominant mode of gene expression evolution in multicellular model organisms. Despite this, they detected evidence for directional selection in their work on primate gene expression, which lead to

human-specific increases in the expression of several transcription factors [22].

However, gene expression might not always be constrained by stabilizing selection alone. Neutral genetic drift can also lead to divergence of gene expression patterns among taxa, making it difficult to distinguish adaptive changes from drift. Wagner [23] showed that increases in mRNA in yeast are also constrained through the energy costs that they incur. Given that yeast have huge effective population sizes (>>1000 individuals), it is predicted that a change in mRNA levels of more than 10% will lead to a decrease in fitness that is sufficient for such a mutation to be effectively removed by purifying selection; thus, these changes in mRNA will not be evolutionarily neutral [23]. In higher organisms, particularly vertebrates, energetic constraints on gene expression might be of minor evolutionary importance because other components of fitness, such as behavior, dominate [23]. In these organisms, which are the focus of most evolutionary microarray studies, larger fitness effects would be required to prevent the fixation of deleterious changes in gene expression. Indeed, Whitehead and Crawford [18] found that a large part of the expression variation could be explained by neutral drift.

It has been suggested that focusing on changes in the mechanisms controlling the global expression profile will reduce the chance of measuring gene expression differences that result from neutral drift. For instance, Rifkin *et al.* [24] reported that the expression of transcription factor (regulatory) genes were less prone to neutral drift than were their downstream targets. Shiu and Borevitz [25] also advocated describing regulatory networks on a genome-wide scale.

Interpreting differential gene expression

There are several technical issues affecting the experimental design of microarray studies in general, and ecological microarray studies in particular. These include the multiple testing problem, caused by the large number of hypotheses tested in parallel; the problem of obtaining a sufficient number of appropriate biological replicates; and the importance of choosing the correct time points for sampling.

Multiple testing problem

The sheer number of measured genes in microarray studies poses important statistical challenges. When testing whether a particular condition, such as temperature or the presence of parasites, has a significant effect on the expression of a single gene, the associated p-value refers only to a single statistical test. However, when testing thousands of genes simultaneously, there is a strong likelihood that some of them will show 'significant' p-values just by chance. This problem is not unique to ecological applications of microarrays. Several approaches for controlling the multiple testing problems have been proposed, the most common of which in microarray analysis is the control of the False Discovery Rate (FDR), introduced by Benjamini and Hochberg [26]. This approach is powerful enough to detect significant effects in multiple testing situations such as microarrays, and is relatively easy to calculate. Other approaches are discussed in Refs [27,28].

Multiple-testing issues can arise in unexpected forms in ecological microarray studies. For example, Bochdanovits and de Jong [29] analyzed a microarray data set (1670 genes) of differentially expressed genes underlying a tradeoff between pre-adult survival and male weight in *D. melanogaster*. At the center of their technique is the evaluation of correlations between gene expression and the two traits. It was argued that it is unlikely that the expression of a gene is simultaneously strongly positively correlated to one trait and strongly negatively correlated to the other trait by chance. Based on this, it was estimated that the fraction of genes that simultaneously occupy the opposite 3.5% tails of the two correlation coefficient distributions is $p = 0.035*0.035*2 = 0.0024$. This would correspond to an expected number of false positives of $E = p*1670 = 4.0$. However, pre-adult survival and male weight were negatively correlated. When this is taken into account, the expected number of false positives is higher, ~31.6 genes per experiment. In this case, the detection of 34 differentially expressed genes, as reported in Ref. [29], corresponds to an expected false discovery rate of FDR = 31.6/34 = 92.9%.

Biological replication

Differential expression estimates also depend on the ability to obtain sufficient independent mRNA replicates from wild populations. Using only technical replication (i.e. repeating the measurements on the same mRNA sample multiple times) can lead to inflated estimates of statistical significance. Hybridizing single samples sufficiently often will necessarily detect statistically significant differences. The calculated p-values will be uninformative in this case. Because p-value calculation is based on the assumption that independent measurements are used, the values are just arbitrary numbers derived from pseudo-replicated data. The results do not provide insight into the causal relationships of differential gene expression. This is important under field conditions, where many unknown factors can affect expression levels. For example, the precise physiological state of the organism at the moment that the sample was collected might vary. Perhaps one of the sampled organisms was well fed, whereas the next sample came from a hungry individual. Another factor could be whether there was any rank in the social hierarchy among individuals. Thus, because gene expression is so sensitive to environmental conditions, animals and plants from field studies should be kept under well-controlled, homogenous laboratory conditions before sampling [30].

Temporal dynamics of gene expression

In addition to environmental and random variation, another confounding problem is the temporal dynamics of gene expression. With some genes only being transiently expressed, one might fail to detect any differences because the timing of measurement is 'wrong'. This became clear in some of the differential expression studies by Voelckel and Baldwin [31]. The authors tried to determine whether plants can distinguish between attacks from insects of different feeding guilds. Indeed, sap-feeding mirids *Tupiocoris notatus* and chewing hornworms *Manduca sexta* elicited different gene expression profiles after an attack that

lasted for 24 h. Yet, after five days, these initial differences had largely subsided and the expression resembled that of plants deserted by the insects [31]. It might therefore be necessary to analyze several time points or developmental stages before reaching a conclusion.

Higher-level analysis

A common approach to analyzing ecological data is the analysis of variance (ANOVA), or its extensions in the form of mixed models. These techniques are straightforward to apply to gene expression experiments (for extensive didactical discussion see Refs [32,33]). A typical model might describe the expression value, y, of a single gene as

$$y = \mu + G + E + T + \varepsilon,$$

where μ is the average expression level, G contains expression differences explained by genetic factors (different lines or populations), E summarizes environmental factors (which can include different tissues), T stands for various kinds on technical factors (e.g. spot, array or batch effects), and ε is the residual variation that is not explained by these other factors.

In an ecological microarray experiment, the term of interest will often be some kind of G×E interaction [e.g. a difference between two populations (G) in their gene expression response to temperature changes (E)]. A recent study applied this approach to expression data from a genetic cross between *Caenorhabditis elegans* lines from Bristol and Hawaii, reared at different temperatures, and identified a large number of genes that show differences in gene expression plasticity (G×E interaction) in the two genetic backgrounds [34].

Although incorporating interactions among multiple factors in the ANOVA model is possible, the interpretation of interaction coefficients at the level of thousands of transcripts is challenging. As an alternative, gene expression responses can be analyzed in the form of Venn diagrams [35], which illustrate how many genes are affected in condition A and condition B, and how many of these overlap (Figure 1). In this case, the influence of each factor is analyzed separately, for example using a *t*-test, and interaction between them is inferred only indirectly. Several Venn diagrams might be necessary to describe the results for up- and downregulated genes, as well as various combinations of these [31,36]. Yet, the necessity to use a fixed threshold for deciding which genes to include in the presented gene sets makes this approach arbitrary.

Breitling *et al.* [37] have introduced a new generalization of the Venn diagram approach, vector analysis, which combines intuitive visualization with a statistical evaluation that helps to detect significant patterns of expression responses in different backgrounds (genetic or environmental) (Figure 2). Such binary visualizations of response difference are a natural way of comparing dynamic expression patterns, and vector analysis provides a statistical basis that makes their interpretation more reliable [37]. This approach has been used, for instance, to compare sex-biased gene expression in two species of *Xenopus*, which have a ZW sex determination system [38]. Using the vector analysis statistics, the authors showed a significant excess of male-biased compared with female-biased genes in both species.

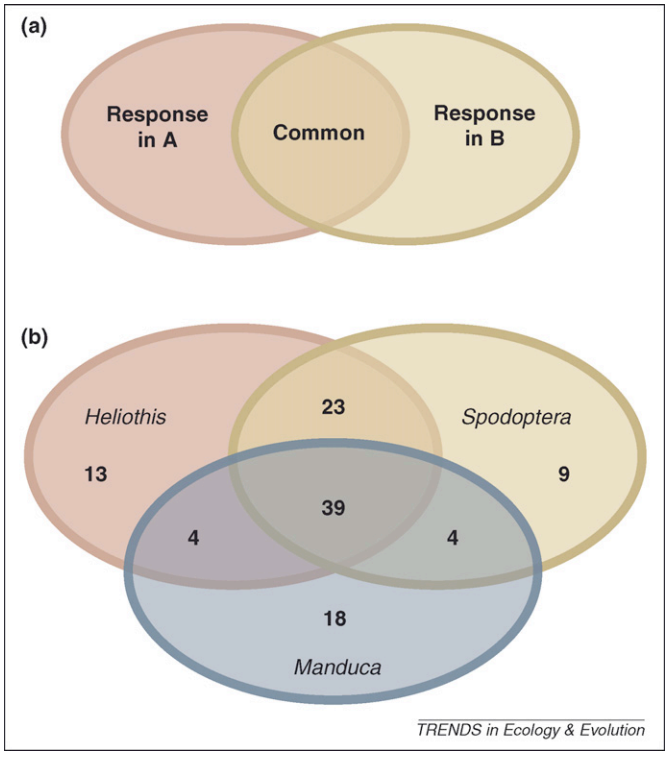

Figure 1. Classical Venn diagram approach to a herbivore attack study. **(a)** This method compares the set of responsive genes in two experimental backgrounds A and B (genetic or environmental). The assignment of genes to the responsive group is usually based on some arbitrary statistical threshold (e.g. a certain significance level or minimum fold-change, which is a measure of the magnitude of a differentially expressed gene). This assignment can be derived for each background independently and does not require a multifactorial analysis. **(b)** An example of Venn diagrams applied to an ecological microarray study of herbivore attack on native tobacco *Nicotiana attenuata* [31]. The number of genes that downregulate expression in response to attack by the larvae of three different moth species (*Manduca sexta, Heliothis virescens* and *Spodoptera exigua*) is shown. There is a large overlap in genes between plants attacked by generalist larvae (*Heliothis* and *Spodoptera*; 23 genes), compared with plants attacked by the specialist *Manduca* (four in each case). Detection of significant patterns of gene expression is facilitated using vector analysis (Figure 2).

Genetical genomics

A more advanced and promising concept of microarrays in ecology involves the detection of genomic loci that underlie variation in gene expression. This approach, called genetical genomics [39], uses microarray data from each individual of a pedigree or experimental cross as a quantitative trait to identify quantitative trait loci (QTL) that influence the expression of genes. Although genetical genomics has already been applied successfully in medical, animal and plant sciences, for instance in studies of tissue specific gene regulation in mice [40–42] and during shoot development in *Arabidopsis* [43], its application to ecology is relatively new. Recently, Street *et al.* [44] used genetical genomics to study the genetics of adaptation to drought in the poplar *Populus*. Using two parental strains with contrasting responses to drought, this study provided candidates for genes responsible for natural variation in drought adaptation. This result could not have been achieved using a common microarray approach where only the two parents are analyzed for gene expression. Li *et al.* [34] showed strong genetic variation of differential expression responses to temperature changes in *C. elegans* and demonstrated the potential of genetical genomics for mapping the molecular determinants of phenotypic plasticity. Although the field of genetical genomics is

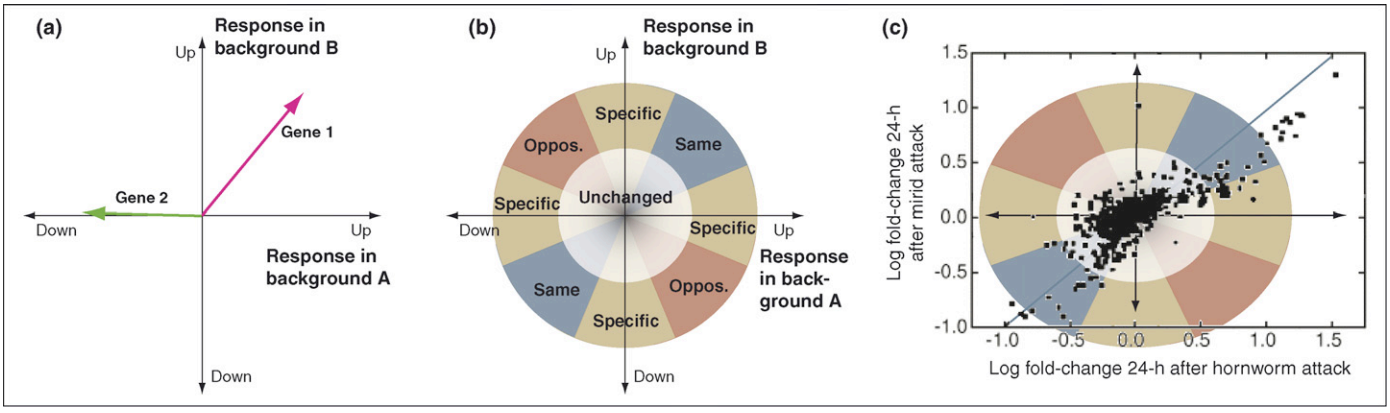

Figure 2. The principle of vector analysis. **(a)** The change in expression of a gene in the two experimental backgrounds is represented by a vector. The two axes correspond to the log-fold changes in the two backgrounds. For example, gene 1 is strongly upregulated in both backgrounds, whereas gene 2 is specifically downregulated in background A, but has lost this response in background B. **(b)** The plane can be systematically subdivided into sectors corresponding to the main behavior types that are possible. In the centre, genes show little, if any, response in either background (white). Other genes respond at similar levels in both backgrounds (blue), are specifically changed in only one background (yellow), or are regulated in opposite directions in background A and B (red). **(c)** Overlaying the vector analysis scheme on gene expression responses in native tobacco *Nicotiana attenuata*, 24 hours after attack by two different herbivore species shows that most genes are either unchanged or show the same response to both attackers (blue sectors). A few genes are specific for one insect species (yellow) and a few show opposite responses (red). The circular shape of the original diagram (b) is transformed into an ellipse, as the axes of the original figure in [69] are not equally scaled. A full vector analysis would assign significance *p*-values to these classifications. (c) reproduced, with permission, from Ref. [69]. To keep the image simple, a few genes in the upper left and lower right quadrants were not included.

still in its infancy, it is envisaged that it will contribute to important discoveries with regard to the genetics of evolutionary trajectories [45].

Confirming microarray data

Because microarray technology is still rapidly developing and many different platform types are used (each with inherent limitations and biases), confirmation experiments are indispensable [46]. Typically, Northern blotting or quantitative real-time PCR (qRT-PCR) are applied to confirm observed expression patterns. Taken literally and applied to all genes, such approaches would sacrifice the two main advantages of microarrays; their rapidity and their genome-wide scope.

Box 1. Biological replication and technical confirmation: a case study

Whitfield *et al*. [8] studied the brain transcriptome underlying individual transitional behavior (nursing or foraging) in the honey bee *Apis mellifera* (Figure I; reproduced with permission from Hans Smid) under field conditions. Their study illustrates the complex interplay of factors that affect ecological array studies and the designs that can deal with this complexity. The following levels of potential confounding factors were controlled:

- Environment: bees were collected from two different host colonies. As behavioral transitions are adjusted to the needs of the hive, this is a crucial factor that could influence observed expression patterns.
- Genetic background: full sisters (75% related owing to haplodiploidy) were compared, using three independent full-sister groups. This technique, which depends on the special sex-determination mechanism of hymenopterans, minimizes genetic variation.
- Age: two different ages (5–9 days and 28–32 days) were considered. As in normal colonies, these age classes generally show only one of the two behavior types (nursing versus foraging); single-cohort colonies were used for obtaining age-matched groups of nurses and foragers.

To obtain statistically useful results, two levels of replication were used: (i) Three individual bees for each combination of factors were collected as biological replicates; and (ii) each sample was hybridized two or four times as technical replicates.

In total, 60 individual brains were profiled. The microarray data were analyzed with both Bayesian statistics and analysis of variance. The most important result was that brain expression differences can predict the behavior of individual bees, based on a few genes. The number of replicates, as well as the control for a large number of environmental factors, enabled the reliability of these predictions to be assessed: each brain sample was withheld from the data set in turn and the remaining samples were used to identify predictor genes, which were then used to predict the phenotype of the withheld sample. This leave-one-out cross-validation showed that, with as little as ten predictor genes, 95% of samples could be assigned to the correct behavioral category.

Confirmation of the observed predictive expression differences again proceeded on several levels. Predictive genes were examined against *Drosophila* genes and assigned to functional categories. This confirmed a few genes that are related to neuronal and behavioral plasticity (e.g. axonogenesis and cell-adhesion genes) or that were identified as related to behavioral transition (i.e. from nursing to foraging). In addition, three selected predictor genes were technically verified by qRT-PCR on independently collected brain samples, using three technical and seven biological replicates (i.e. 21 data points per gene). Differences in mRNA levels were in the predicted direction for all three genes compared with the microarray hybridization experiment.

Figure I.

Currently, there is no consensus on the most efficient targeted confirmation strategy [47]. This is evident from the few ecological microarray papers that have technically confirmed the microarray data (Table 1): they all used different selection criteria. Lai *et al.* [36] arbitrarily picked a few cDNA ESTs for RT-PCR that did or did not show differential expression; Wang *et al.* [5] complemented their array analysis with a Northern blot hybridization to validate a single differentially expressed gene of special interest; and Juenger *et al.* [48] selected five candidate genes that were differentially expressed, and then performed RT-PCR. Moreover, just repeating the expression measurements with a different quantitative technique is not informative considering the environmental (rather than technical) confounders discussed earlier. Biological validation is required (i.e. confirming that the differentially expressed genes are causally related to the studied phenomenon). A study that combines technical confirmation and biological validation in a convincing fashion is discussed in Box 1.

Ultimately, the challenge is to prove that the identified genes are important for acclimation or adaptation. A promising strategy would be to make use of the standing genetic variation in wild populations, for example by extensive comparative genotyping of transcriptional control regions identified in genetical genomics experiments [34]. The task would be to test whether the predicted genetic differences do correlate with differences in adaptive capacity. However, such an approach will essentially remain correlational. For causal validation, it will therefore be necessary to use molecular and genetic manipulations, such as gene knock-outs or overexpression in an ecological setting. However, such studies are currently still under development.

Recommendations and future challenge

In conclusion, we suggest that researchers should be aware of the sensitivity of gene expression levels to environmental variation and should consider carefully whether the populations under study are likely to show evolutionarily constrained differences in gene expression. Furthermore, we advise that multiple testing is corrected for by controlling the false discovery rate and that sufficient biological replicates are obtained from the field to achieve statistical significance.

Once differential expression profiles have been reliably determined, the question 'what do these differences mean ecologically?' remains. At present, microarrays are typically used to identify genes that are differentially expressed between environments and/or genotypes. This is largely a descriptive approach that is relatively devoid of *a priori* hypotheses. However, phenotypic traits are increasingly being investigated in detail at the molecular level, and the insight into regulatory gene pathways is rapidly expanding [45]. For instance, the signaling cascades leading to the initiation of flowering in plants are understood in molecular detail [49], although refinements and additional details are frequently published. In such a case, it will become possible to proceed from gene expression observations to targeted interventions such as gene knock-outs [50] and controlled overexpression [51]. Only then will ecological microarray experiments convert from being a largely descriptive approach to a more hypothesis-driven experimental science.

Acknowledgements

We thank J. Coolon, A. Doroszuk, A. van der Wurff, W. van der Putten, M. Bezemer, Z. Bochdanovits, C. Voelckel, N. Aubin-Horth, D. Crawford and five anonymous referees for constructive comments.

References

1 Adams, M.D. *et al.* (2000) The genome sequence of *Drosophila melanogaster*. *Science* 287, 2185–2195
2 The Arabidopsis initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815
3 Goffeau, A. *et al.* (1996) Life with 6000 genes. *Science* 274, 546–567
4 Derome, N. *et al.* (2006) Parallelism in gene transcription among sympatric lake whitefish (*Coregonus clupeaformis* Mitchill) ecotypes. *Mol. Ecol.* 15, 1239–1249
5 Wang, Z. *et al.* (2006) A cDNA macroarray approach to parasite-induced gene expression changes in a songbird host: genetic response of house finches to experimental infection by *Mycoplasma gallisepticum*. *Mol. Ecol.* 15, 1263–1273
6 Ralph, S. *et al.* (2006) Genomics of hybrid poplar (*Populus trichocarpa* × *deltoides*) interacting with forest tent caterpillars (*Malacosoma disstria*): normalized and full-length cDNA libraries, expressed sequence tags, and a cDNA microarray for the study of insect-induced defences in poplar. *Mol. Ecol.* 15, 1275–1297
7 Aubin-Horth, N. *et al.* (2005) Interaction of rearing environment and reproductive tactic on gene expression profiles in Atlantic salmon. *J. Hered.* 96, 261–278
8 Whitfield, C.W. *et al.* (2003) Gene expression profiles in the brain predict behavior in individual honey bees. *Science* 302, 296–299
9 Whitehead, A. and Crawford, D.L. (2006) Variation within and among species in gene expression: raw material for evolution. *Mol. Ecol.* 15, 1197–1211
10 Shields, R. (2006) MIAME we have a problem. *Trends Genet.* 22, 65–66
11 Carsten, L.D. *et al.* (2005) Gene expression patterns accompanying a dietary shift in *Drosophila melanogaster*. *Mol. Ecol.* 14, 3203–3208
12 Seltmann, M. *et al.* (2005) Assessment of a systematic expression profiling approach in ENU-induced mouse mutant lines. *Mamm. Genome* 16, 1–10
13 Hughes, T.R. *et al.* (2000) Functional discovery via a compendium of expression profiles. *Cell* 102, 109–126
14 DeRisi, J.L. *et al.* (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278, 680–686
15 Matsubara, S. *et al.* (2005) Nocturnal changes in leaf growth of *Populus deltoides* are controlled by cytoplasmic growth. *Planta* 223, 1315–1328
16 Ju, Z. *et al.* (2002) Differential gene expression in the brain of channel catfish (*Ictalurus punctatus*) in response to cold acclimation. *Mol. Genet. Genom.* 268, 87–95
17 Casati, P. and Walbot, V. (2003) Gene expression profiling in response to ultraviolet radiation in maize genotypes with varying flavonoid content. *Plant Physiol.* 132, 1739–1754
18 Whitehead, A. and Crawford, D.L. (2006) Neutral and adaptive variation in gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 103, 5425–5430
19 Ranz, J.M. and Machado, C.A. (2003) Uncovering evolutionary patterns of gene expression using microarrays. *Trends Ecol. Evol.* 21, 29–37
20 Wray, G.A. *et al.* (2003) The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* 20, 1377–1419
21 Gilad, Y. *et al.* (2006) Natural selection on gene expression. *Trends Genet.* 22, 456–461
22 Gilad, Y. *et al.* (2006) Expression profiling in primates reveals a rapid evolution of human transcription factors. *Nature* 440, 242–245
23 Wagner, A. (2005) Energy constraints on the evolution of gene expression. *Mol. Biol. Evol.* 22, 1365–1374
24 Rifkin, S.A. *et al.* (2003) Evolution of gene expression in the *Drosophila melanogaster* subgroup. *Nat. Genet.* 33, 138–144
25 Shiu, S.-H. and Borevitz, J.O. (2006) The next generation of microarray research: applications in evolutionary and ecological genomics. *Heredity* DOI: 10.1038/sj.hdy.6800916
26 Benjamini, Y. and Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57, 289–300

27 Breitling, R. (2006) Biological microarray interpretation: the rules of engagement. *Biochem. Biophys. Acta* 1759, 319–327
28 Nettleton, D. (2006) A discussion of statistical methods for design and analysis of microarray experiments for plant scientists. *Plant Cell* 18, 2112–2121
29 Bochdanovits, Z. and De Jong, G. (2004) Antagonistic pleiotropy for life-history traits at the gene expression level. *Proc. R. Soc. B* 271 (Suppl. 3), , pp. S75–S78
30 Oleksiak, M.F. *et al.* (2002) Variation in gene expression within and among natural populations. *Nat. Genet.* 32, 261–266
31 Voelckel, C. and Baldwin, I.T. (2004) Generalist and specialist lepidopteran larvae elicit different transcriptional responses in *Nicotiana attenuata*, which correlate with larval FAC profiles. *Ecol. Lett.* 7, 770–775
32 Kerr, M.K. and Churchill, G.A. (2001) Statistical design and the analysis of gene expression microarray data. *Genet. Res.* 77, 123–128
33 Rosa, G.J.M. *et al.* (2005) Reassessing design and analysis of two-colour microarray experiments using mixed effects models. *Comp. Funct. Genom.* 6, 123–131
34 Li, Y. *et al.* (2007) Mapping determinants of gene expression plasticity by genetical genomics in *C. elegans*. *PLoS Genet.* 2, e222 DOI: 10.1371/journal.pgen.0020222
35 Kestler, H.A. *et al.* (2005) Generalized Venn diagrams: a new method of visualizing complex genetic set relations. *Bioinformatics* 21, 1592–1595
36 Lai *et al.* (2006) Microarray analysis reveals differential gene expression in hybrid sunflower species. *Mol. Ecol.* 15, 1213–1228
37 Breitling, R. *et al.* (2004) Vector analysis as a fast and easy method to compare gene expression responses between different experimental backgrounds. *BMC Bioinf.* 6, 181
38 Malone, J.H. *et al.* (2006) Sex-biased gene expression in a ZW sex determination system. *J. Mol. Evol.* 63, 427–436
39 Jansen, R.C. and Nap, J.-P. (2001) Genetical genomics: the added value from segregation. *Trends Genet.* 17, 388–391
40 Cotsapas, C.J. *et al.* (2006) Genetic dissection of gene regulation in multiple mouse tissues. *Mammal. Genome* 17, 490–495
41 Chesler, E.J. *et al.* (2005) Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks that modulate nervous system function. *Nat. Genet.* 37, 233–242
42 Bystrykh, L. *et al.* (2005) Uncovering regulatory pathways that affect hematopoietic stem cell function using 'genetical genomics'. *Nat. Genet.* 37, 225–232
43 DeCook, R. *et al.* (2006) Genetic regulation of gene expression during shoot development in *Arabidopsis*. *Genetics* 172, 1155–1164
44 Street, N.R. (2006) The genetics and genomics of the drought response in *Populus*. *Plant J.* 48, 321–341
45 Rockman, M.V. and Kruglyak, L. (2006) Genetics of global gene expression. *Nat. Rev. Genet.* 7, 862–872
46 Rockett, J.C. and Hellmann, G.M. (2004) Confirming microarray data – is it really necessary? *Genomics* 83, 541–549
47 Allison, D.B. *et al.* (2006) Microarray data analysis: from disarray to consolidation and consensus. *Nat. Rev. Genet.* 7, 55–65
48 Juenger, T.E. *et al.* (2006) Natural genetic variation in whole-genome expression in *Arabidopsis thaliana*: the impact of physiological QTL introgression. *Mol. Ecol.* 15, 1351–1366
49 Blazquez, M.A. (2000) Flower development pathways. *J. Cell Science* 113, 3547–3548
50 Brown, D.M. *et al.* (2005) Identification of novel genes in *Arabidopsis* involved in secondary cell wall formation using expression profiling and reverse genetics. *Plant Cell* 17, 2281–2295
51 Zhuang, X. *et al.* (2006) Over-expression of OsAGAP, an ARF-GAP, interferes with auxin influx, vesicle trafficking and root development. *Plant J.* 48, 581–591
52 Quéré Le, A. *et al.* (2006) Screening for rapidly evolving genes in the ectomycorrhizal fungus *Paxillus involutus* using cDNA microarrays. *Mol. Ecol.* 15, 535–550
53 Quéré Le, A. *et al.* (2004) Divergence in gene expression related to variation in host specificity of an ectomycorrhizal fungus. *Mol. Ecol.* 13, 3809–3819
54 Townsend, J.P. *et al.* (2003) Population genetic variation in genome-wide gene expression. *Mol. Biol. Evol.* 20, 955–963
55 Ferea, T. *et al.* (1999) Systematic changes in gene expression patterns following adaptive evolution in yeast. *Proc. Natl. Acad. Sci. USA.* 96, 9721–9726
56 Schenk, P.M. *et al.* (2000) Coordinated plant defense responses in *Arabidopsis* revealed by microarray analysis. *Proc. Natl. Acad. Sci. USA* 97, 11655–11660
57 Filatov, V. *et al.* (2006) Comparison of gene expression in segregating families identifies genes and genomic regions involved in a novel adaptation, zinc hyperaccumulation. *Mol. Ecol.* 15, 3045–3059
58 Kreps, J.A. *et al.* (2002) Transcriptome changes for *Arabidopsis* in response to salt, osmotic, and cold stress. *Plant Physiol.* 130, 2129–2141
59 Hegarty, M.J. *et al.* (2005) Development of anonymous cDNA microarrays to study changes to the *Senecio* floral transcriptome during hybrid speciation. *Mol. Ecol.* 14, 2493–2510
60 Roda *et al.* (2004) Individual variability in herbivore-specific elicitors from the plant's perspective. *Mol. Ecol.* 13, 2421–2433
61 Halitschke, R. *et al.* (2003) Molecular interactions between the specialist herbivore *Manduca sexta* (Lepidoptera, Sphingidae) and its natural host *Nicotiana attenuata*. *VI.* Microarray analysis reveals that most herbivore-specific transcriptional changes are mediated by fatty acid-amino acid conjugates. *Plant Physiol.* 131, 1894–1902
62 Schmidt *et al.* (2004) *Solanum nigrum*: a model ecological expression system and its tools. *Mol. Ecol.* 13, 981–995
63 Wertheim, B. *et al.* (2005) Genome-wide gene expression in response to parasitoid attack in *Drosophila*. *Genome Biol.* 6, R94
64 Bochdanovits, Z. *et al.* (2003) Covariation of larval gene expression and adult body size in natural populations of *Drosophila melanogaster*. *Mol. Biol. Evol.* 20, 1760–1766
65 Grozinger, C.M. *et al.* (2003) Pheromone-mediated gene expression in the honey bee brain. *Proc. Natl. Acad. Sci. U. S. A.* 100, 14519–14525
66 Podrabsky, J.E. and Somero, G.N. (2004) Changes in gene expression associated with acclimation to constant temperatures and fluctuating daily temperatures in an annual killifish *Austrofundulus limnaeus*. *J. Exp. Biol.* 207, 2237–2254
67 Oleksiak, M.F. *et al.* (2005) Natural variation in cardiac metabolism and gene expression in *Fundulus heteroclitus*. *Nat. Genet.* 37, 67–72
68 Whitehead, J.A. and Crawford, D.L. (2005) Variation in tissue-specific gene expression among natural populations. *Genome Biol.* 6, R13
69 Voelckel, C. *et al.* (2004) Herbivore-induced plant vaccination. Part II. Array-studies reveal the transience of herbivore-specific transcriptional imprints and a distinct imprint from stress combinations. *Plant J.* 38, 650–663